

Structure and Method for Forming the Gate Electrode in a Multiple-Gate Transistor

TECHNICAL FIELD

[0001] The present invention relates to semiconductor devices and the preferred embodiment of the present invention provides an improved structure and method for forming the gate electrode in a multiple-gate transistor.

BACKGROUND

[0002] The dominant semiconductor technology used for the manufacture of ultra-large scale integrated (ULSI) circuits is the metal-oxide-semiconductor field effect transistor (MOSFET) technology. Reduction in the size of MOSFETs has provided continued improvement in speed performance, circuit density, and cost per unit function over the past few decades. As the gate length of the conventional bulk MOSFET is reduced, the source and drain increasingly interact with the channel and gain influence on the channel potential. Consequently, a transistor with a short gate length suffers from problems related to the inability of the gate to substantially control the on and off states of the channel.

[0003] Phenomena such as reduced gate control associated with transistors with short channel lengths are termed short-channel effects. Increased body doping concentration, reduced gate oxide thickness, and ultra-shallow source/drain junctions are ways to suppress short-channel effects. However, for device scaling well into the sub-50 nm regime, the requirements for body-doping concentration, gate oxide thickness, and source/drain (S/D) doping profiles become increasingly difficult to meet when conventional device structures based on bulk silicon (Si)

substrates are employed. Innovations in front-end process technologies or the introduction of alternative device structures will sustain the historical pace of device scaling.

[0004] For device scaling well into the sub-30-nm regime, a promising approach to controlling short-channel effects is to use an alternative transistor structure with more than one gate, i.e. multiple-gates. An example of the alternative transistor structure is the multiple-gate transistor. A multiple-gate transistor 100 has a plan view as shown in Figure 1.

[0005] Referring to Figure 1, the transistor 100 includes a silicon fin 102 overlying an insulator layer 104 over a silicon substrate 114 (see Figures 2a, 2b or 2c). A gate dielectric (not explicitly shown) covers a portion of the silicon fin 102. A gate electrode 106 straddles across the silicon fin 102. The gate dielectric isolates the gate electrode 106 from the silicon fin 102.

[0006] Examples of the multiple-gate transistor include the double-gate transistor (as shown in Patent No. 6,391,695, and X. Huang *et al.*, "Sub-50 nm p-channel finFET," *IEEE Trans. Electron Devices*, vol. 48, no. 5, pp. 880-886, May 2001), triple-gate transistor (as shown in R. Chau *et al.*, "Advanced depleted-substrate transistors: single-gate, double-gate, and tri-gate", 2002 International Conference on Solid State Devices and Materials, Nagoya, Japan, pp. 68-69, Sep. 2002), omega field-effect transistor (FET) (as shown in F.-L. Yang *et al.*, "25 nm CMOS Omega-FETs," *International Electron Device Meeting*, Dig. Technical Papers, Dec. 2002), and the surround-gate or wrap-around gate transistor (as shown in J. P. Colinge *et al.*, "Silicon-on-insulator gate-all-around device," *International Electron Device Meeting*, Dig. Technical Papers, pp. 595-598, Dec. 1990 and E. Leobandung *et al.*, "Wire-channel and wrap-around-gate metal-oxide-semiconductor field-effect transistors with a significant reduction of short channel effects," *J. Vacuum Science and Technology B*, vol. 15, no. 6, pp. 2791-2794, 1997). Each of these references is incorporated herein by reference.

[0007] A multiple-gate transistor structure is expected to extend the scalability of CMOS technology beyond the limitations of the conventional bulk MOSFET and realize the ultimate limit of silicon MOSFETs. The introduction of additional gates improves the capacitance coupling between the gates and the channel, increases the control of the channel potential by the gate, helps suppress short channel effects, and prolongs the scalability of the MOS transistor.

[0008] The simplest example of a multiple-gate transistor is the double-gate transistor, as described in U.S. Patent No. 6,413,802 issued to Hu, *et al.* U.S. Patent No. 6,413,802 is incorporated herein by reference. As illustrated in a cross-sectional view in Figure 2a, the double-gate transistor 100 has a gate electrode 106 that straddles across the channel within the fin-like silicon body 102, thus forming a double-gate structure. There are two gates, one on each sidewall 108 of the silicon fin 102, and separated from the fin 102 by gate dielectric 100. An etchant mask 112 overlies a top surface of fin 102. The plan view of the double-gate structure is shown in Figure 1.

[0009] In U.S. Patent No. 6,413,802, the transistor channel comprises a thin silicon fin defined using an etchant mask and formed on an insulator layer, e.g. silicon oxide. Gate oxidation is performed, followed by gate deposition and gate patterning to form a double-gate structure overlying the sides of the fin. Both the source-to-drain direction and the gate-to-gate direction are in the plane of the substrate surface.

[0010] Another example of the multiple-gate transistor is the triple-gate transistor. A cross-section of the triple-gate transistor 100 is illustrated in Figure 2b and thus the plan view of the triple-gate structure is the same as the double gate structure shown in Figure 1. The triple-gate transistor structure 100 has a gate electrode 106 that forms three gates: one gate on the top surface 116 of the silicon body/fin 102, and two gates on the sidewalls 108 of the silicon

body/fin 102. The triple-gate transistor achieves better gate control than the double-gate transistor because of it has one more gate on the top of the silicon fin.

[0011] The triple-gate transistor structure may be modified for improved gate control, as illustrated in Figure 2c. Such a structure 100 is also known as the Omega (Ω) field-effect transistor (FET), or simply omega-FET, since the gate electrode 106 has an omega-shape in its cross-sectional view. The encroachment of the gate electrode 106 under the semiconductor fin or body 102 forms an omega-shaped gate structure. This encroachment results in notch or undercut region 132 as shown in Figure 2c. It closely resembles the Gate-All-Around (GAA) transistor for excellent scalability, and uses a very manufacturable process similar to that of the double-gate or triple-gate transistor.

[0012] The omega-FET has a top gate, adjacent surface 110, two sidewall gates, adjacent sidewalls 108, and special gate extensions or encroachments 118 under the fin-like semiconductor body 102. The omega-FET is therefore a field effect transistor with a gate 106 that almost wraps around the body. In fact, the longer the gate extension, i.e., the greater the extent of the encroachment E , the more the structure approaches or resembles the gate-all-around structure. The encroachment of the gate electrode 106 under the silicon body 102 helps to shield the channel from electric field lines from the drain and improves gate-to-channel controllability, thus alleviating the drain-induced barrier lowering effect and improving short-channel performance.

[0013] The multiple-gate transistor structures described, e.g., the double-gate transistor, the triple-gate transistor, and the omega-FET, have a common feature: a gate electrode 106 that straddles across the fin-like semiconductor active region 102. The formation of the gate electrode 106 involves a definition step using techniques such as photolithography, and an

etching step. The formation of the gate electrode 106 over a large step height introduced by the semiconductor fin 102 presents a very challenging problem. For example, if the top surface of the gate electrode 106 is not substantially flat, the patterning of the gate electrode using lithographic methods can be difficult due to focusing problems.

[0014] Figure 3 shows a prior art process for forming a gate electrode 106 in a multiple-gate transistor. In Figure 3a, a gate electrode material 120 is deposited over a semiconductor fin 102 and covered with a gate dielectric 104. As shown, the top surface of the gate electrode material 120 is non-planar due to the fin 102.

[0015] A mask material 122 such as a photoresist is then deposited on the gate electrode material 120, as shown in Figure 3b. The top surface of the mask material is usually a planar surface. As a result, the thickness of the mask material varies from t_1 in one region to t_2 in another region.

[0016] A lithographic mask 124 that includes an opaque region 126 and a transparent region 128 will be used to pattern the mask material 122. The pattern on the lithographic mask 124, however, may not be accurately transferred to the mask material 122 due to the varying thickness of the mask material 122. As a result, the patterned mask material 130 may be formed with different widths, as shown in Figure 3c. When the gate electrode material 120 is subsequently etched, as shown in Figures 3d and 3e, the gate electrode 106 may be formed with a non-uniform gate length. The uniformity of the critical gate length dimension is therefore adversely affected.

[0017] The present invention provides simple and improved methods for the formation of the gate electrode in a multiple-gate transistor.

SUMMARY OF THE INVENTION

[0018] These and other problems are generally solved or circumvented, and technical advantages are generally achieved, by preferred embodiments of the present invention which provide a structure and method for forming the gate electrode in a multiple-gate transistor.

[0019] In accordance with a preferred embodiment of the present invention, a method for forming a gate electrode of a multiple-gate transistor includes providing a semiconductor structure that includes a semiconductor fin overlying an insulator layer. The structure also includes a gate dielectric overlying at least a portion of the semiconductor fin. First and second gate electrode material are formed over the gate dielectric. A patterned mask is formed over a second gate electrode material, and the first and second gate electrode materials are etched.

[0020] In accordance with another preferred embodiment of the present invention, a structure for a multiple-gate transistor includes a semiconductor fin overlying an insulator layer. A gate dielectric overlies at least a portion of the semiconductor fin. A gate electrode overlies the gate dielectric and includes a first gate electrode material underlying a second gate electrode material. Source and drain regions are formed in portions of the semiconductor fin oppositely adjacent to said gate electrode.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] For a more complete understanding of the present invention, and the advantages thereof, reference is now made to the following descriptions taken in conjunction with the accompanying drawing, in which:

[0022] Figure 1 is a plan view of a multiple gate transistor;

[0023] Figure 2a is a cross-sectional view of the double-gate transistor;

[0024] Figure 2b is a cross-sectional view of the triple-gate transistor;

[0025] Figure 2c is a triple-gate structure with recessed insulator;

[0026] Figures 3a-3e illustrate a conventional process flow for forming a multiple gate transistor;

[0027] Figure 4 is a flow chart showing a method of forming gate electrode;

[0028] Figures 5a-5i illustrate a process flow of the present invention for forming a multiple gate transistor;

[0029] Figure 6 is a perspective view showing formation of spacers;

[0030] Figure 7 is a perspective view showing epitaxial source and drain regions; and

[0031] Figure 8 is a perspective view showing the case where the CMP process exposes the first gate electrode material.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0032] The making and using of the presently preferred embodiments are discussed in detail below. It should be appreciated, however, that the present invention provides many applicable inventive concepts that can be embodied in a wide variety of specific contexts. The specific embodiments discussed are merely illustrative of specific ways to make and use the invention, and do not limit the scope of the invention.

[0033] The preferred embodiments of the present invention provide several improved methods for the formation of semiconductor devices and the resulting structures. These embodiments will now be described in the context of the formation of the gate electrode in a multiple-gate transistor.

[0034] Multiple-gate transistors such as the double-gate transistor, the triple-gate transistor, the omega-FET, have a common feature: a gate electrode that straddles across a semiconductor fin-like active region. Such devices are also known as semiconductor fin devices or FinFETs. The semiconductor fin has a predetermined fin height h and a predetermined fin width w . Another common feature of multiple-gate transistors is that the sidewall surfaces of the semiconductor fins are used for current conduction. A significant amount of source-to-drain current in a typical multiple-gate transistor is carried along the sidewall surfaces. Essentially, the effective device width of the multiple-gate transistor is a function of the fin height h . The larger the fin height, the large amount of current the device can deliver. However, when the fin height is large, the formation of the gate electrode becomes very challenging. The preferred embodiment of this invention teaches a structure and method for forming the gate electrode of the multiple-gate transistor and may be better understood in the following embodiments.

[0035] A method of forming a gate electrode of the multiple-gate transistor is illustrated using the flow chart of Figure 4. Three-dimensional perspectives of the multiple-gate transistor during the various process steps described in Figure 4 are illustrated in Figures 5a-5i. The formation of device 200 begins with a semiconductor-on-insulator substrate that includes a semiconductor layer 202 overlying an insulator layer 204, as shown in Figure 5a. The insulator layer 204 overlies a substrate 206. The semiconductor layer 202 may be formed from an elemental semiconductor such as silicon, an alloy semiconductor such as silicon-germanium, or a compound semiconductor such as gallium arsenide or indium phosphide. The semiconductor layer 206 is preferably silicon. The thickness of the semiconductor layer may be in the range of about 200 angstroms to about 5000 angstroms. In an alternate embodiment, bulk semiconductor substrates such as a bulk silicon substrate may also be used.

[0036] The insulator layer 204 may be formed from any dielectric or insulator, and is preferably comprised of silicon oxide or silicon nitride or a structured combination of both. The insulator layer 204 may have a thickness in the range of about 100 angstroms to about 2000 angstroms, although it is understood that thinner or thicker thicknesses may be used. The substrate 206 may be a silicon substrate, for example.

[0037] Referring now to Figure 5b, at least one semiconductor fin 208 is formed by patterning the semiconductor layer 202. The semiconductor fin patterning process may be accomplished by depositing a commonly used mask material (not shown) such as photoresist or silicon oxide over the semiconductor layer 202. The mask material is then patterned and the semiconductor layer is etched in accordance with the pattern. In this manner, a semiconductor structure including at least one semiconductor fin overlying an insulator layer is formed.

[0038] A gate dielectric layer 210 is then formed on the semiconductor fin 208, as shown in Figure 5c. The gate dielectric 210 may be formed by thermal oxidation, chemical vapor deposition, sputtering, or any other methods known and used in the art for forming a gate dielectric. Depending on the technique of gate dielectric formation, the gate dielectric 210 thickness on the top of the fin 208 may be different from the gate dielectric thickness on the fin sidewall. In one embodiment, the gate dielectric thickness on the top surface of the fin is less than 20 angstroms.

[0039] The gate dielectric 210 may be formed from a material such as silicon dioxide or silicon oxynitride with a thickness ranging from about 3 angstroms to about 100 angstroms, preferably less than about 10 angstroms. The gate dielectric 210 may alternatively be formed from a high permittivity (high-k) material (e.g., with a relative permittivity greater than about 5) such as lanthanum oxide (La_2O_3), aluminum oxide (Al_2O_3), hafnium oxide (HfO_2), hafnium oxynitride HfON, or zirconium oxide (ZrO_2), or combinations thereof, with an equivalent oxide thickness of 3 angstroms to 100 angstroms.

[0040] Next, as shown in Figure 5d, gate electrode material 212 is deposited. In this embodiment, the gate electrode material 212 includes a first gate electrode material 214 underlying a second gate electrode material 216. In other embodiments, additional gate electrode materials may be included. According to this embodiment, the first and second gate electrode materials 214 and 216 have substantially different etch rates in a gate electrode etching process. The first and second gate electrode materials are conductive materials.

[0041] The first gate electrode material 214 acts as an etch-stop layer for the etching of the second gate electrode material 216. In the preferred embodiment, the first and second gate electrode materials may be selected from polycrystalline-silicon (poly-Si), poly-crystalline

silicon-germanium (poly-SiGe), metallic nitrides, metallic silicides, metallic oxides, and metals. Examples of metallic nitrides include tungsten nitride, molybdenum nitride, titanium nitride, and tantalum nitride, or their combinations. Examples of metallic silicide include tungsten silicide, titanium silicide, cobalt silicide, nickel silicide, platinum silicide, erbium silicide, or their combinations. Examples of metallic oxides include ruthenium oxide, indium tin oxide, or their combinations. Examples of metal include tungsten, titanium, aluminum, copper, molybdenum, nickel, platinum, and others.

[0042] The first and second gate electrode materials 214 and 216 may be deposited by chemical vapor deposition (CVD), by sputter deposition, or by other techniques known and used in the art for depositing conductive materials. The thickness of the first or the second gate electrode material may be in the range of about 200 angstroms to about 4000 angstroms.

[0043] In one embodiment, the first gate electrode material 214 is a metallic nitride, and the second gate electrode material 216 is selected from polycrystalline-silicon (poly-Si), polycrystalline silicon-germanium (poly-SiGe), metallic nitrides, metallic silicides, metallic oxides, and metals. For example, the first gate electrode material 214 is tantalum nitride, and the second gate electrode material 216 is nickel silicide. In another embodiment, the first gate electrode material 214 is a metal, and the second gate electrode material 216 is selected from poly-Si, poly-SiGe, metallic nitrides, metallic silicides, metallic oxides, and metals.

[0044] After deposition, the top surface of the second gate electrode material 216 usually has a non-planar top surface, as shown in Figure 5d. The non-planar surface may have a step height that is proportional (sometimes approximately equal) to the fin height 218. The step height can introduce significant problems in the transfer of a predefined pattern from a lithographic mask onto the gate electrode material 212. Ions may or may not be introduced into

the gate electrode material 212 at this point. Ions may be introduced, for example, by ion implantation techniques. The ion implantation introduces the dopants such that a peak dopant concentration occurs at an implant depth below the non-planar top surface.

[0045] Referring next to Figure 5e, a chemical mechanical polishing (CMP) process is performed to planarize the top surface of the second gate electrode material 216. The CMP process may or may not expose the first gate electrode material 214. In Figure 5e, the CMP process does not expose first gate electrode material 214. Figure 8 is provided to show the case where the first gate electrode material 214 is exposed. The root-mean-square surface roughness of the planarized top surface of the gate electrode material 216 is preferably less than 100 angstroms.

[0046] The definition of gate electrode 220 will now be described with respect to Figures 5e-5g. A mask material 222 is formed on the substantially planarized top surface of the gate electrode material 216, as shown in Figure 5f. The mask material 222 is then patterned to create a patterned mask 224 as shown in Figure 5g. As a result of the substantially planarized top surface, a desired predefined pattern can be accurately transferred onto the mask material 222 to form a patterned mask 224.

[0047] Referring to Figure 5h, the pattern is transferred to the second gate electrode material 216 by an etching process, preferably a dry plasma etching process. Since the second gate electrode material 216 and the first gate electrode material 214 have substantially different etch rates, the etching of the second gate electrode material can stop on the first gate electrode material. The resulting structure is shown in Figure 5h.

[0048] The first gate electrode material 214 may then be etched, stopping on the insulator layer 204. The etching of the first gate electrode material 214 may be accomplished using a dry

plasma etching process or a wet etching process. For example, if the first gate electrode material 214 is titanium nitride, the dry plasma etching process can be a plasma etching process employing fluorine chemistry, and the wet etching process may employ an etchant comprising ammonia hydroxide and hydrogen peroxide. Such a wet etchant etches the titanium nitride first gate electrode material with a high selectivity with respect to the underlying silicon oxide insulator layer.

[0049] The patterned mask 224 may then be removed as shown in Figure 5i. The portion of the gate dielectric 210 not covered by the gate electrode 212 may or may not be removed during the etching process. In the case where some gate dielectric remains on the semiconductor fin 208 not covered by the gate electrode 212, the gate dielectric 210 may be subsequently removed by dry or wet etching.

[0050] The process of Figures 5a-5i illustrated the formation of a triple-gate transistor device. It is understood that similar process steps could be used to form any other multiple gate transistor. For example, an etchant mask (see element 112 in Figure 2a) can be formed over the semiconductor fin 208. Similarly, the insulating layer 204 can be recessed, resulting in a notch (see element 132 of Figure 2c) at the base of the semiconductor fin 208. Any of the other features incorporated in Figures 2a, 2b and 2c could similarly be incorporated in a device of the present invention.

[0051] Referring now to Figure 6, the source and drain regions may then be formed. The formation of the source and drain regions may involve many steps. In the preferred embodiment, an ion implantation process is first performed to dope the source and drain regions 228 and 230 immediately adjacent to the channel region. The channel region is the portion of the semiconductor fin 208 wrapped around by the gate dielectric 210 and the gate electrode 212.

[0052] Spacers 226 are then formed on the sidewalls of the gate electrode 212. The spacers 226 may be formed by deposition of a spacer material followed by anisotropic etching of the spacer material. The spacer material comprises of a dielectric material, preferably silicon nitride but alternately silicon oxide or another insulating material. The spacer material may also be comprised of a stack of dielectric materials, such as a silicon nitride layer overlying a silicon oxide layer.

[0053] In the preferred embodiment, a selective epitaxy is performed to increase the width and height of the fin in the source and drain regions 228 and 230. The selective epitaxy results in epitaxial growth in the source and drain regions, and perhaps the gate electrode region. Epitaxial layer 232 shown in Figure 7 illustrates this feature.

[0054] An optional ion implantation is then performed to dope the source and drain regions 228 and 230. For example, the ion implantation process can be performed if the selective epitaxy does not incorporate dopants into the grown regions during epitaxial growth.

[0055] Figures 9a and 9b illustrate an alternate embodiment of the present invention. Figure 9a begins with the structure of Figure 5d. In other words, each of the steps described above with respect to Figures 5a-5d can be performed in the alternate embodiment. In the alternate embodiment, however, only one conductive gate electrode material is necessary. As a result, Figure 9a shows only a single gate electrode material 214. It is understood, however, that two or more materials can be used.

[0056] Referring now to Figure 9a, a planarizing material 234 is formed above gate electrode material 214. Planarizing material 234 is preferably a dielectric but any material can be used. Preferably, planarizing material 234 should have an etch rate that is substantially the same as the etch rate of gate electrode material 214 for a given etching process. The planarizing

material 234 should also be a material that can be deposited in a substantially planar manner (e.g., smooth within about 100 angstroms peak to valley). In the preferred embodiment, the planarizing material is chemical-vapor-deposited silicon-rich oxide when the gate electrode material is polysilicon. In other embodiments, the planarizing material can be spin-on glass, silicon oxide, or doped glass (e.g., BPSG, PSG, or FSG).

[0057] Referring now to Figure 9b, the planarizing material 234 and the gate electrode material 214 are etched to create a substantially planar top surface. Preferably, the entire layer of planarizing material 234 is removed. The etching process can be an etch back using a fluorine or chlorine etch chemistry with a gate electrode material of polysilicon and a planarizing layer of silicon-rich oxide. Alternatively, a chemical mechanical polish step can be used.

[0058] While several embodiments of the invention, together with modifications thereof, have been described in detail herein and illustrated in the accompanying drawings, it will be evident that various modifications are possible without departing from the scope of the present invention. The examples given are intended to be illustrative rather than exclusive. The drawings may not necessarily be to scale and features may be shown in a schematic form.